

The NIST Chemistry WebBook

The NIST Chemistry WebBook is a site on the Internet (<http://webbook.nist.gov/>) that distributes physical and chemical property data on well defined chemical species and reactions. The purpose of this activity is to improve the nation's science and technology efforts by providing convenient access to NIST Standard Reference Data^{®‡} to scientists, engineers, and the general public.

P.J. Linstrom and W.G. Mallard (Div. 838)

The data in the WebBook is from databases developed by NIST and outside contributors. Such data were traditionally published only in hardcopy format. With the widespread introduction of personal computers SRD products were distributed in the form of databases on floppy disks and CD-ROMs. The development of the Internet provided a third avenue for distribution of NIST data. The NIST Chemistry WebBook was one of the first efforts to distribute this data over the Internet. This has several advantages for NIST's customers: rapid access to data from any location with Internet access, updates and corrections made without any action required by the user, and better integration of data and relevant documentation. The NIST Chemistry WebBook is a service, so it is available on a continuing basis. Users worldwide can (and do) access the site 24 hours a day.

The greatest technical challenge in this effort is the integration of databases from multiple sources into a single collection of data. Because of the lack of standards for physical and chemical properties, the data collections included in the site use diverse formats and conventions for reporting data. Problems with integrating these data sets are exacerbated by traditional chemical informatics problems associated with the identification of chemical species: inconsistent nomenclature, ambiguous or erroneous references to 3rd party registries, and inconsistent conventions for chemical formulas. Systematically resolving these issues provides a common infrastructure for searching, identifying, and providing reliable information. Additionally, the WebBook employs internationally recognized developments in nomenclature, data exchange standards, and chemical identifiers, such as InChI - see next section.

The site was established in 1996 and has grown to encompass a wide variety of thermochemical, ion energetics, physical, solubility, spectroscopic, and chromatographic

data. The site also includes two special features: a set of interactive physical property models developed at the NIST Boulder labs and a group additivity based estimator for gas phase thermodynamic properties. The physical property models provide thermodynamic and transport property data at user supplied conditions for a number of industrially important fluids. The group-additivity based property starts with user supplied structures using Benson increments.

The NIST Chemistry WebBook is part of the NIST Standard Reference Data program and is the result of the work of many contributors. These are identified in the "credits" section of the web page at

<http://webbook.nist.gov/chemistry/>.

The site is regularly updated with new functionality and new and expanded data set. The history of major changes in the site can be found at

<http://webbook.nist.gov/chemistry/history.html>.

As a result of work sponsored by the European Union and efforts at NIST, the site is now available in Spanish, Portuguese, French, and Czech.

These translations are invisible to users that primarily speak English; they only show up for users with web browsers configured to prefer a language other than English.

In addition to displaying data, the NIST Chemistry WebBook also provides links other, relevant, on-line databases at NIST. This is done at the chemical species level, so the WebBook is effectively an index of the other sites. Linked databases include the gas phase kinetics database, the computational chemistry comparison and benchmark database, the NIST atomic spectra database, the NIST microwave spectra database, the electron-impact ionization cross section database, and the thermophysical properties of hydrogen web site.

Many of the recent improvements in the site have focused on the use of the IUPAC International Identifier (InChI) to improve interoperability with other data resources. The site has supported display and searching based on the identifier since shortly after its release. As illustrated in figure 1, InChI is used by the site to identify related chemical species such as optical and geometric isomers. In the past year the search algorithm used by the site was enhanced to bet-

[‡] cf. United States Congress in the Standard Reference Data Act (United States Code, title 15, chapter 7A, section 290).

ter handle identifiers which are generated with a fixed hydrogen layer. This was done to improve interoperability with PubChem web site operated by National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH). The IUPAC identifier is used to link species in the NCBI site to the NIST Chemistry WebBook.

Figure 1: Portion of web page showing display and use of the IUPAC International Chemical Identifier

Other significant work carried out in the past year included updates to the electronic spectra, chemical structure, and retention index databases.

The NIST Chemistry WebBook is the most widely used SRD product developed by NIST.

It is used by scientists and engineers in industry and academia and by students at the high school, college and graduate school levels. As noted in figure 2, usage of the site has grown over time. The site now regularly receives more than two million page views per month.

Alanine

- **Formula:** $C_3H_7NO_2$
- **Molecular weight:** 89.09
- **IUPAC International Chemical Identifier:**

[O-]C(=O)C([NH3+])C([O-])=O
[Download the identifier in a file.](#)

- **CAS Registry Number:** 56 41 7
- **Chemical structure:**



The IUPAC International Chemical Identifier is generated from the chemical structure.

This structure is also available as a [2d Mol file](#) or as a [computed 3d Mol file](#).

- **Stereoisomers:**

• [DL-Alanine](#)
 • [D-Alanine](#)

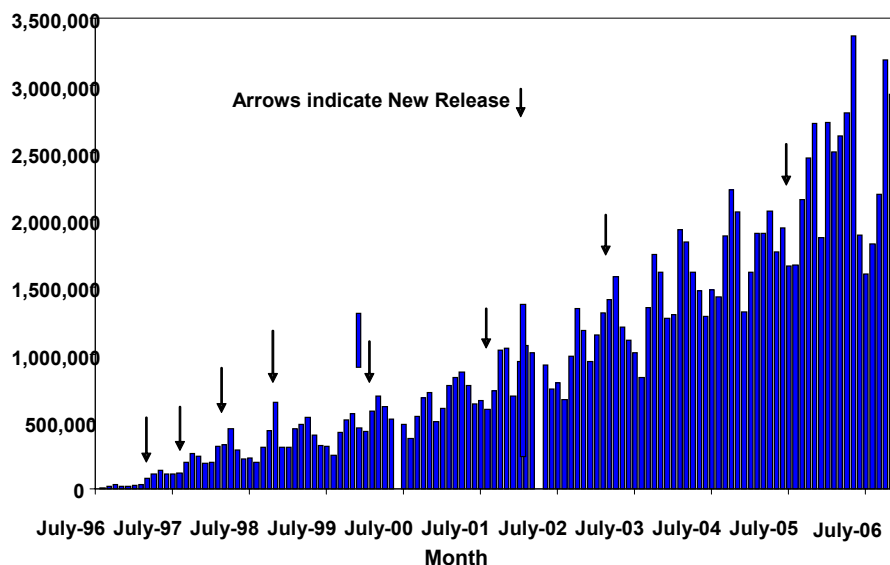
} These related species are found using the IUPAC International Chemical Identifier

Future Plans:

This is an ongoing activity. Enhancements to the site's data collections and functionality occur continually over time. The following items are expected for FY 2007:

1. Updates to the ion and ion-cluster thermochemical databases.
2. Improved support cases where a search for a species based on its InChI fails. This change should also make it easier for others to link to specific species in the site.
3. Infrastructure enhancements to better support the storage and display of data for chemical reactions.

Figure 2: Usage history of the NIST Chemistry WebBook in page views per month.



Publications:

The site itself is a NIST publication:
 P.J. Linstrom and W.G. Mallard,
 Eds., NIST Chemistry WebBook,
 NIST Standard Reference Data-
 base Number 69, June 2005, Na-
 tional Institute of Standards and
 Technology, Gaithersburg MD,
 20899 <http://webbook.nist.gov>